**Bioinformatics Exercise 5:** Pairwise sequence alignment

Pairwise sequence alignment methods are concerned, in contrast to BLAST, with finding the best-matching piecewise local or global alignments of protein and DNA sequences.

*Local alignment (Smith-Waterman algorithm)*

Local alignment methods find related regions within sequences - in other words they can consist of a subset of the characters within each sequence (e.g. positions 20-40 of sequence A might align with positions 50-70 of sequence B).This is a more flexible technique than global alignment and has the advantage that related regions which appear in a different order in the two proteins (which is known as domain shuffling) can be identified as being related.

*Global alignment (Needleman-Wunsch algorithm)*

A global alignment between two sequences is an alignment in which all of the characters in both sequences participate in the alignment. Global alignments are useful mostly for finding closely-related sequences.

**EMBOSS-Align**

The EMBOSS-Align tool contains two programs each using a different algorithm:

* When you want an alignment that covers the whole length of both sequences, use the needle program
* When you are trying to find the best region of similarity between two sequences, use the water program

Open [EMBOSS](https://www.ebi.ac.uk/Tools/emboss/) and then select the “protein” option under “Water” (the local alignment)

* Compare the protein sequences for the Human dystrophin protein with that of the chicken dystrophin protein. You can also try this again with the mouse and the fruit fly sequences.
* Repeat this process but use “Water” (the global alignment)
* **Which alignment is better?**

Note that matching amino acids are connected with a "|" symbol. Mismatches would be connected with a space. A gap would be represented with a "-" symbol. Similar amino acids (e.g. threonine vs methionine) are connected via a "." symbol. Thus a sequence alignment can be represented in the format...

DMFCNTEGQGIAMM

| |||| ..

TMG--NEGQGSETT